

QUCCOO – QUery ConstruCtion with OntOlogy, ontology-based search interface

The Department of Information Studies at the Tampere University educates professionals for service, administrative, planning and research tasks in the whole sector of information management, including library and information services. The information management market both in the private and public sectors demands professionals with a sound knowledge provided by Information Studies. The aspects of Information Studies particularly emphasized in the curriculum are information retrieval, information seeking and information management.

Abstract

Ontologies and other conceptual models describe the structure of delimited topical areas. While conceptual models are traditional tools in Information Science, e.g., in the form of thesauri, they have acquired much recent attention in research in several disciplines due to the semantic web. While ontologies greatly resemble traditional thesauri, they can be richer in structural relationships. The most important difference is however the aim toward computational semantics (or inference), which may support more “intelligent” applications and interoperability of IR systems. Through the use of ontologies, the information searcher can avoid (at least greatly reduce) the complexity of natural languages when searching, e.g., in the Web. Annotations based on ontologies are supposed to capture the semantic content of documents in a nutshell. As the history of indexing research informs, however, there are problems in cost, quality (consistency), exhaustiveness and specificity of annotation. For example, the most popular metadata format for the Web, the Dublin Core format, was in 2002 employed in 0.3 % of web documents.

Our approach is therefore different: we investigate ontology-based access to unannotated document collections. This line of research was begun some 20 years ago and has produced several academic degrees and research articles. We have shown that structured queries, based on ontologies, greatly improve performance in ontology-based IR at least in news article collections.

Quccoo, based on the 3-level architecture, provides direct access to different database contents as well as the Web with the help of ontologies. Quccoo shows the following characteristics:

- Engine independence:
The Quccoo interface provides uniform access to various kinds of full text database engines with their proprietary query languages.
- Indexing independence:
Collections may be indexed with or without morphological normalization, e.g. lemmatization, stemming or neither, without the user having to care about this.
- Collection independence:
Different kinds of collection can be searched because they need not be annotated.

- Expression independence:
Natural language expressions in document texts are varied. The Quccoo interface accommodates synonyms of selected concepts and varied word order, and thus abolishes the burden of searching for various expression forms of the concept.
- Expression variability:
In agglunative languages, expressions could also show diverse forms as tokens. The Quccoo interface takes care of these.
- Language independence:
Queries can be formulated in several languages with the help of Quccoo interface easily.

From the software architecture point of view, Quccoo is a web application, which uses state-of-art servlet technology; it supports diverse full-text database engines, e.g. Trip, Inquiry, Lemur, etc. as well as web search engines like Google. Its intuitive interface simplifies reaching the documents. It also supports multilingual and cross-lingual search. The motto, "Just select your concepts, targets, and go!" describes the aim of Quccoo in short.

Researchers:

Mr. Feza Baskaya – supervisor Prof. Kalervo Järvelin

Ms. Anne Kakkonen – supervisor Prof. Jaana Kekäläinen